Volcano Detection without Ground Truth

Padhraic Smyth*, Usama Fayyad
Jet Prepulsion Laboratory MS 238-420,
California Institute of Technology, 4800 Oak Grove Drive Pasadena, CA 91109

Michael Burl, Pietro Perona
EE Dept., California Institute of Technology 116-81, Pasadena, CA 91125

Abstract

Manual detection and characterization of small geological objects in remotely sensed imagery can be a time-consuming and tedious task for planetary geologists. For example, in the recently acquired 30,000 synthetic-aperture radar images from the Magellan mission to Venus, there are estimated to be on the order of 10^5 to 10^6 mall volcanoes visible in the data [1]. In this paper we describe the development of a pattern recognition system which aims to automate the low-level task of volcano detection. In particular, we focus on the problems posed by the lack of absolute ground-truth.

The overall pattern recognition system consists of focus of *attention* via *matched* filtering, feature generation from local regions of interest (ROIs) via the use of singular-value decomposition filters, and subsequent discrimination of ROIs into volcano and non-volcano categories via standard classification models [2]. In order to provide training data for the system, planetary geologists familiar with Venusian volcanism identify the likely locations of volcanoes in particular training images. The geologists assign to tack such candidate volcano a predetermined level of certainty, which is in turn mapped to a subjectively determined posterior probability value. Quantization helps calibration by anchoring the subjective estimates to specific predefined qualitative categories. The probabilistic estimates are useful in two respects. Firstly they can be used during the training/estimation process (as described [3]): fractions of training exemplars are assigned to each class according to the subjective probability estimate associated with that class. Secondly, the probabilistic estimates can be directly used for performance assessment of both algorithms and humans relative to the reference consensus (we use consensus-derived estimates as our reference and use individually-derived estimates to calibrate individual scientist performance). Probabilistic labels lead naturally to the notion of *probabilistic ROC* curves - we describe their use aud interpretation. While different geologists operate at different points along the overall curve in terms of absolute discrimination (corresponding to different utility functions), their estimated probabilistic ROC performance is quite similar. Empirical results thus far indicate that the current algorithmic system is close to the performance of geologists on high resolution images in terms of detection, but is relatively weak in terms of its ability to provide accurate probabilistic estimates. We discuss the strengths and weaknesses of the current approach and describe ongoing work and future directions of the research.

References

1. J. E. Guest et al., 'Small volcanic edifices in the plains of Venus,' *Journal of Geophysical Research*, vol. 97, no. E10, pp. 15949 15966, October 25, 1992.
2. M. C. Burl et al., 'Automating the hunt for volcanoes on Venus,' CNS Technical Report 34, Computation and Neural Systems Dept., California Institute of Technology, December 15th, 1993.
3. P. Smyth, 'Learning with probabilistic supervision,' in *Computational Learning Theory and Natural Learning Systems 3*, T. Petsche, M. Kearns, S. Hanson, R. Rivest (eds), Cambridge, MA: MIT Press, in press.

Topic: **Applications** Presentation: **Poster**

*Author to whom correspondence should be addressed, *pjs@galway.jpl.nasa.gov*. The research described in this paper was performed at the Jet Prepulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration and was supported in part by ARPA under grant number N00014-92-J-1860.